DOCUMENT RESUME

ED 326 048 FL 018 966

AUTHOR Stansfield, Charles W.
TITLE IDEA Oral Language Proficiency Test (IPT II).
PUB DATE 90
NOTE 18p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classification; *English (Second Language); Language Proficiency; *Language Tests; Listening Comprehension; *Oral Language; Screening Tests; Secondary Education; Speech Skills; *Test Format; Test Reliability; *Test Use; Test Validity
IDENTIFIERS *IDEA Oral Language Proficiency Test

ABSTRACT

The IDEA Oral Language Proficiency Test (IPT II), an individually-administered measure of speaking and listening proficiency in English as a Second Language designed for secondary school students, is described and discussed. The test consists of 91 items and requires 5-25 minutes to administer. Raw scores are converted to one of seven proficiency level scores, which are in turn used to classify the student as non-English-speaking (NES), limited-English-speaking (LES), and fluent-English-speaking (FES). The test's format and evolution are outlined in the context of the overall IDEA Oral Language Proficiency Test series. Construction, pilot testing, materials, items, and administration procedures are included in the discussion. A section on the practical applications of the test examines its use: (1) as a pre- and post-test; (2) as a measure of language dominance when used in conjunction with the Spanish version; and (3) for diagnosis of student strengths and weaknesses. Finally, studies of the test's validity and reliability as originally reportd in a professional manual are discussed, and it is concluded that the test scored high on both measures, and that the only significant weakness was the unclear manner in which the studies were reported. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED326048

Charles W. Stansfield, Ph.D.
Director, ERIC Clearinghouse for Languages and Linguistics;
Director, Division of Foreign Language Education and Testing,
Center for Applied Linguistics, Washington, DC.

IDEA ORAL LANGUAGE PROFICIENCY TEST (IPT II)

Enrique F. Dalton and Beverly A. Amori. Brea, California: Ballard & Tighe, Inc.

Introduction

Designed for students in grades 7-12, the IDEA Oral Language Proficiency Test (IPT II) is an individually-administered measure of speaking and listening proficiency in English as a second language (ESL). The test contains 91 items and requires between 5 and 25 minutes to administer, depending on the student's level of proficiency. The average administration time is 15 minutes. Raw scores are converted to one of seven proficiency level scores. The proficiency level score is, in turn, used to classify the student as non-English-speaking (NES), limited English-speaking (LES), or fluent English-speaking (FES). The IPT II is a part of the IDEA Oral Language Proficiency Test series. The series includes a Pre-IPT in English and Spanish for pre-kindergarten children, an IPT I English and Spanish for grades K-6, and the IPT II in English and Spanish for grades 7-12. This review focuses on

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G R Tucker
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL 018966

the IPT II in English.

Since the IPT II is part of the IDEA Oral Language Proficiency Test series, its history is best described in the context of that series, whose history begins in the early 1970s. At that time, two public elementary school teachers, Wanda Ballard and Phyllis Tighe, were teaching in the Los Angeles area. During a six year period, these two teachars developed a set of oral language development materials for their students. The success of these materials, called Individualized Developmental English Activities (IDEA), led to their publication in 1976. The following year, a parallel set of materials, Ideas para el desarrollo del espanol por actividades, was developed in Spanish by Dr. Enrique F. Dalton. A natural consequence of the development of the oral language program was the development of a proficiency test that could be used to place students in the IDEA program or in others. This process began in 1978; Forms A and B of the IPT I in English were published in the fall of 1979. The validation studies for this test were directed by Dr. Dalton. He also played the lead role in the development of the IPT I in Spanish, which was published late in 1980, and wrote the Technical Manuals for all the IPT tests. After these tests were completed, work began on the IPT II, which was published in September, 1983. A description of the development of the IPT II follows.

In May 1982, a Committee of Language Specialists consisting of seven experienced teachers of ESL and bilingual education, and specialists in oral language development in California was formed

and met to advise the authors on the development of a comprehensive list of oral English language skills important at the secondary level. The authors, Enrique Dalton and Beveraly Amori, then began the process of developing such a list, which was refined in subsequent meetings of the committee. At least four items for each skill on the list were written by the authors. These items were then ranked according to their suitability and quality by each committee member. Over 300 items were written by the authors. Some of these items were based on the oral language skills contained in the eight levels of the IDEA program, and others based on research in second language development, including basic interpersonal communicative skills (BICS) and cognitive/academic language proficiency (CALP) (Cummins, 1984). Each item was also ranked according to the seven proficiency levels on the IPT scale.

Following deletion of items deemed inappropriate or repetitive, two parallel forms of a pilot test were developed and administered to a small group (number not indicated in the Technical Manual) of monolingual English-speaking students during December, 1982. Using item difficulty and discrimination indices from this pilot testing, revisions were made on the items themselves and in the sequencing of items. Subsequently, a field testing of each form was conducted during the Spring of 1983. This field testing involved 306 monolingual native English-speaking students in grades 7-12, as well as an additional 153 students in those grades who were classified as non-English speaking (NES), limited English speaking (LES), and fluent English speaking (FES).

A total of 120 of the 306 monolingual English-speaking students were retested in order to determine parallel form and interrater reliability. The data from this field testing is the basis for the reliability and validity information in the Technical Manual.

The IPT II consists of a set of materials that sells for $92.00 (1990 price). Each set contains either 50 student test booklets or 50 diagnostic score cards. The component desired must be stipulated at the time the set is ordered. The set also contains a book of 15 stimulus pictures, an Examiner's Manual, a Technical Manual, 50 proficiency test level summaries, which describe the skills that normally have been acquired by students at each level, and 10 group lists, which show students' test level scores and NES/LES/FES classifications.

The test consists of a series of questions or instructions to the student. Most items (93%) require an oral response. The remaining 7% of the items test comprehension by requiring the student to make some physical response such as pointing to something in a stimulus picture. These five comprehension items focus on vocabulary while testing parts of the body, spatial relations, time, ordinal numbers, superlatives.

Most of the oral production items test vocabulary either through a question/answer or a sentence completion format, with the response based on one of the stimulus pictures (Exp. "What is this?" or "We cook soup on the ..."). The vocabulary tested relates to the school, geometric shapes, pet animals, days of the week, and the vegetables that make up a salad at the lower levels,

to coins, holidays and a number of adjectives at the upper levels. Other oral production items test syntax, often through a question/answer format, also based on the stimulus pictures (Exp. "Where's she going?" ..To the movies.) In other cases, a descriptive prelude provides background information that is used to shape a desired response involving syntax (Exp. "Mr. Lee had a book about horses. His brother wanted to read it. What did Mr. Lee do with the book?"). Some items test syntax through a yes/no question format (Exp. "Do you know how to fly a helicopter?"). In the latter example, the student is told to answer in a complete sentence. Some items use questions and picture stimuli to test morphology (Exp. "Whose sweater is this? ..Its hers."). The critical feature being tested is the /s/ morpheme of third person singular feminine possessive pronoun. At the highest levels, the test also taps an organizational/expressive ability by asking the examinee to complete a story and to retell in his or her own words a story read by the examiner.

Practical Applications/Uses

The IPT II can be used to assess the oral language proficiency of students in grades 7-12. Some confusion exists as to whether the IPT I and, by extension, the IPT II are tests of general proficiency or achievement tests oriented to a specific set of instructional materials. The issue centers on the fact that the IPT I can be used to place students within the eight levels that make up the IDEA Oral Language Program, which is designed for use

in grades K through 8. While this debate over the nature of the test may be logical for the IPT I, it is inappropriate to extend it to the IPT II. The IDEA Oral Language Program does not extend beyond grade 8 and no claims are made in the IPT II Examiner's Manual or the Technical Manual that the test can be used for placement within another set of IDEA instructional materials.

Due to its length, the IPT is sensitive to gains in overall language proficiency. Therefore, the two forms of the test can be used as pre and post test measures to identify gains in language skills. The identification of such gains is often a desirable part of the evaluation of a special instructional program, such as an ESL program, migrant education, bilingual education, or compensatory education.

The IPT II can also be used, jointly with the IPT II in Spanish to determine language dominance; that is, the language in which the student is most proficient. To do this, first, the IPT II-English level score is used to classify the student as NES, LES, or FES. Next, the level score in the child's native language is used to classify the child's proficiency in the home language in a similar manner. Thus, a child might be classified as non-Spanish speaking (NSS), limited Spanish speaking (LSS), or fluent Spanish speaking (FSS). Finally, and if necessary, the two classifications can be compared to place the child in one of the five Lau language dominance categories (Office of Civil Rights, 1975).

The IPT II, like most tests, can also be used to diagnose a student's strengths and weaknesses. Diagnostic Score Cards (DSCs)

can be ordered instead of the test booklets for this purpose. The DSC links each item to a matrix of skills assessed by the test (vocabulary, morphology, syntax, comprehension, as discussed at the end of the previous section). This matrix is similar to a test "blueprint," which is often used to demonstrate a content validity. When using the DSC, the examiner reads the questions from the test booklet, but records the response on the DSC. The DSC is then placed in the student's cumulative folder.

The IPT is administered to one student at a time. The authors recommend that the examiner be bilingual in English and the language of the student. Either English or the student's native language can be used to explain the test procedures prior to the start of the test. Following 4 sample items, the examiner begins with the first 14 items, which are associated with level score A. These items test very basic vocabulary. At the end of the section, the student's performance is scored. A student making four or more errors is given level score A and the test is discontinued. If 3 or fewer errors are made, the students is asked the 15 questions associated with level score B. A poor performance on this part (8 or more errors) will again place the student at level score A. If the student makes 4 to 7 errors, the student is given score level B and the test is discontinued. If a student makes 3 or fewer errors, the examiner proceeds to ask the 15 questions associated with level score C. The test continues in similar fashion through the last part which contains the 16 questions associated with level score F. Thus, on any given part, the student may earn a score

that either a) places him or her at the previous level, b) places him or her at the current level, or c) advances him or her to the next part. Students who answer 75% of the items in level F correctly, are assigned a level score of M, meaning mastery of the skills assessed on the test.

The examiner points to one of the IPT II Test Pictures on 31 of the 91 questions. Depending on the student's response, the examiner places a check mark in the box labeled "Correct" or "Incorrect" in the student test booklet. To aid the examiner in scoring, the test booklet lists a critical feature of each response that must be present in order for the response to be marked correct. When there is more than one possible correct response, the alternatives are indicated with a slash mark (/). If the response calls for a complete sentence, the examiner cues the student "Answer in a sentence." Or, the examiner may say the first part of the sentence and wait for the student to continue the response and provide the critical feature. Since the IPT is scored in a relatively objective, straightforward manner, examiners can usually learn or be trained to administer and score it in half a day or less.

The time required to administer the IPT averages about 15 minutes, and varies between 5 and 25 minutes according to the number of items that are presented to the student. This, in turn, may vary according to the student's proficiency. More proficient students are presented with more parts and more items. However, if an examiner has prior knowledge that a student has some ability

in English, the examiner may skip the items associated with the lower level scores and proceed directly to the middle level scores, thereby reducing the total administration time. In such cases, if the student misses more that one of the first six items on a given level, the examiner should descend to the previous level and begin again. At the end of the test, the examiner uses the level score attained by the student to assign an NES/LES/FES classification based on a chart on the back of the student's test booklet.

Technical Aspects

Several studies were conducted by the authors in order to address the validity of the IPT II. However, the way they are reported in the Technical Manual is neither clear, organized, or logical, and sometimes inappropriate subjects were used in these studies. As indicated above, these studies were conducted in the Spring of 1983.

The first studies involved 186 of the 306 monolingual English-speaking students who participated in the field testing. These students' English teachers were asked to predict the IPT II level scores of their students based on the list of oral language skills associated with each score level. The list is printed on the IDEA Proficiency Test Summary which is part of the test package. The predicted score level of these students was then correlated with the attained score level. The correlations for both forms were low and not significant.

This should not be surprising for two reasons. First, since

most English teachers do not emphasize instruction in oral language skills, they would not be prepared to make accurate judgements about their students' oral language skills. Indeed, they would probably base such judgements on their students' writing ability, which is what is emphasized in the secondary school English curricula. Second, since 93% of these native English speakers scored at levels F or M, there were few differences in their scores. Without differentiation in scores, there is no possibility of correlation. Yet this latter explanation is not mentioned in the Technical Manual. Finally, it seems inappropriate to correlate predicted with attained scores for a sample of native English speakers. The IPT II is a test for ESL learners, and such tests are, by definition, not designed for the native English-speaking population. Thus, in spite of the fact that this low correlation was needlessly included in the Technical Manual, it's lack of significance should not be a source of concern.

Similar observations can be made regarding the efforts reported in the Technical Manual to correlate IPT results with CTBS scores, age, grade, writing proficiency, math proficiency, etcetera, of this sample of native English speakers. None of these correlations were significant and it is not clear why this data was gathered or why it is presented in the Technical Manual.

One useful outcome of the above study on native English speakers was that it corroborated the designation of levels F and M as the Fluent English Speaking (FES) classification. Thus, nonnative English speakers who attain these levels can be said to

score at the native English speaker level on the test.

Fortunately, a second study was conducted during the Spring of 1983 involving 153 nonnative speakers of English. 78 of these students took Form A while 75 took Form B. Again, the Technical Manual reports the results of a correlation analysis with student age and grade for this sample. Not surprisingly, the IPT was found not to correlate with age or grade. Of course, there is no reason why English proficiency should correlate with age or grade for a sample of nonnative English speakers. An 18 year old immigrant who has just arrived in the U.S. will usually have far less proficiency than a 12 year old who has been in the U.S. for three years. Thus, it would be more reasonable to expect English proficiency to correlate with the amount of time that each subject had been in the United States. This, in fact, is what the Committee of Language Specialists recommended, with the result that additional data on time in country was gathered from student files. For a sample of 99 students, the correlation between IPT level score and time in country was found to be .62. Among this group, 49 took Form A while 50 took Form B, and the correlation for each group was almost identical. This provides some meaningful evidence of the validity of the IPT II.

The English teachers of the same group of 153 nonnative English speaking students were asked to predict the IPT II level scores of their students based on the list of oral language skills associated with each score level. The list is printed on the IDEA Proficiency Test Summary which is part of the test package. The

predicted score level of these students was then correlated with the attained score level. The correlation (.66 and .43) for both forms was significant. This again provides some meaningful evidence of the validity of the IPT.

The Technical Manual reports that the IPT scores of the same sample of 153 nonnatives were compared with the FLS/LES/NES classifications previously determined by the school district. These FES/LES/NES classifications were obtained using three other tests approved for use in California by the California Department of Education. These tests were the Language Assessment Battery, the Language Assessment Scales, and the Bilingual Syntax Measure. The correlation with district classification was found to be .56 for Form A and .36 for Form B. While both correlations were significant, it is not clear why Form B did not perform as well.

Finally, the IPT II scores of this sample were compared with the FES/LES/NES classifications made by teachers on the basis of their knowledge of the students' oral language ability, academic ability, and other unobtrusive measures. The correlation was .68 for Form A and .59 for Form B. Both these correlations are significant also.

An important validity issue is the method used to determine what constitutes an NES/LES/FES classification. In this case, the authors compared teacher and district classifications of 148 nonnative English speaking students with their IPT II level scores. The results of this comparison were used to determine the IPT score levels that correspond to each classification. For the IPT II,

score level A corresponds to a classification as non-English speaking (NES). Score levels B through E correspond to classification as limited English speaking (LES). And similarly, score levels F and M correspond to a classification as fluent English speaking (FES). This latter correspondence agrees with the results of the first study of native English speakers reported earlier.

Two studies of the reliability of the IPT I are reported in the Technical Manual. In the Spring of 1983, the 153 students mentioned earlier took one form of the IPT II. 78 took Form A and 75 took Form B. An analysis of the internal consistency reliability (Cronbach's Alpha) showed that the internal consistency reliability of Forms A and B was .98. This is exceptionally high reliability for any test, and especially for a productive skills test.

Test/retest reliability was determined in the following manner. A sample of 30 monolingual English speaking students was administered Form A by different examiners at one week intervals. The correlation between the scores on the two different administrations was .43. This low correlation was due to the fact that little variance was found among the group on either administration. 22 of the 30 students attained level score M on both administrations and 29 attained either E or M on both administration. Again, this was due to the fact that an English-only sample was selected for this study. The study should have been conducted on nonnative rather than native speakers of English,

since the test was designed to discriminate among nonnative speakers. A similar study involving 30 students who took Form B twice found a test/retest reliability of .73. Although this correlation is higher than that found for Form A, it is probably well below the true test/retest reliability that would be attained with an appropriate sample of nonnative English speakers.

In another study, which attempted to assess parallel form reliability when different raters are used, 56 monolingual English speaking students were administered both forms of the test, each by a different rater, within a one week interval. This approach takes into account error in measurement attributable to both different forms and different raters. The resulting correlation, .24, was not significant. Had the same rater been used, the parallel-form reliability would probably have been slightly higher. However, the principal cause of this low correlation was the fact that an inappropriate sample was selected. Had the sample been composed of learners of English as a second language, undoubtedly the reliability coefficient would have been much higher. In theory, the parallel form reliability should approximate internal consistency reliability, which was found to be .98 for samples of nonnative English speakers.

Critique

The IPT II was developed by practicing teachers with many years of classroom experience. The combination of their experience and the test's length have ensured that the test has adequate

content validity. The content validity is outlined in a blueprint for each form in the Technical Manual. The IPT II is also easy to administer and score, and the 15 minutes average administration is not excessive for an individually administered test, except perhaps for large districts with intake centers that need to assess thousands of students within a few days at the beginning of each school year. The system for converting level scores to language proficiency classifications appears sound.

Several validity studies show that the test correlates well with teacher ratings of language proficiency and with teacher's classifications into an NES/LES/FES category for nonnative English-speaking students. There is also evidence of its relationship to school achievement. The reliability is also high, perhaps due to its length, the similarity of the two forms, and the relative ease with which one can learn to score it accurately.

Only a couple of weaknesses can be identified in the test. The major weakness seems to be the Technical Manual. The research reported in it is not described clearly. The Manual contains many tables but little narrative explanation. As a result, it is difficult for a test user to put these tables together in order to arrive at a more complete understanding of the test's reliability and validity. Rivera and Zeller (1987) noted the same problem with the manual in their review of the IPT II. Since use of the IPT II is increasing, especially in California and Texas, the publisher should consider producing a new manual that would present the development and validation of the test in a clear manner.

A second problem is that inappropriate samples, consisting of native English speakers, were used to present evidence of reliability and validity. The result of this error was a failure to demonstrate adequate reliability or validity when such samples were involved. The test publisher should consider conducting further studies using samples of nonnative English speakers and then reporting the data in a revised Technical Manual. Correlations with other relevant data, such as scores on other ESL proficiency tests and scores on standardized achievement tests could then be presented for nonnative English speakers, thereby providing a more comprehensive and meaningful analysis of the instrument. Given the large number of users of this test, it should not be difficult to collect such data.

Although it may be somewhat premature to say so, given the dirth of relevant empirical research, this reviewer tends to agree with the publisher's claim that the test can be used as a test of overall oral language proficiency for students in grades 7-12.

For another review of this test, see Rivera and Zeller (1987).

References

Office of Civil Rights. (1975). Task force findings specifying remedies available for eliminating past educational practices ruled unlawful under Lau versus Nichols. Federal Register, Summer. Also, Washington, D.C.: U.S. Department of Health, Education, and Welfare.

Cummins, J. (1984). Wanted: A theoretical perspective for relating language proficiency to academic achievement among bilingual students. In C. Rivera (Ed.), Language proficiency and academic achievement (pp. 2-19). Clevedon, Avon, England: Multilingual Matters.

Rivera, C. & Zeller, A. M. (1987). [Review of the Idea Proficiency Test II]. In J.C. Alderson, K.J. Krahnke, & C.W. Stansfield (Eds.), Reviews of English language proficiency tests (pp. 39-41). Washington, DC: Teachers of English to Speakers of Other Languages.